

---

〈論 文〉

## 正準相関分析による家計の所得と消費の関係\*

山 田 智 哉

---

### 1 はじめに

消費支出は平成5年に景気低迷を受けて減少となって以降、実質減少が続いている。平成15年の実質増減率も-0.8%となっている。この原因として考えられるのは景気低迷による所得の減少によるものが大きい。このように消費は所得に応じて決定付けられるとして、主に消費と所得の関係を決定する消費関数が経済学で重要な役割をしている。

消費関数を考える場合、いくつかの所得仮説がある。一般に知られているのはケインズの絶対所得仮説である。これは絶対所得が大きくなれば、消費も比例して大きくなるという仮説で、 $C$ 、 $Y$  をそれぞれ所得、収入とすると、消費関数は

$$C = C_0 + cY$$

で与えられる。ここで  $c$  は限界消費性向、 $C_0$  は最低水準の消費をあらわす。その後、消費予測に失敗したことなどにより、恒常所得仮説あるいはライフサイクル仮説が考えられた。これは消費が現在及び将来の予想所得の加重平均である「恒常所得」に依存するという仮説である(林(1986))。また家計行動の理論モデルを考える場合、家計の効用関数や所得、物価、資産といった変数により所得、消費、貯蓄といった関係をモデル化したものが考えられる。特に消費関数に関しては所得処分モデルから関係式を得ることができる。さらに近年では消費に基づく資本資産価格モデル(C-CAPM)による消費推計(羽森(1996))も行われている。

これらの消費関数推計を考える場合、主に消費としては1財のみを考える場合が多く、家計調査における消費支出が用いられる。またこれらの消費推計は消費の絶対額に対する推計であることから、消費の配分などは考慮に入れない。ライフサイクル所得仮説やC-CAPMは時系列的な変動分を評価するため、今期以外の消費や所得についても考えるが、これも同様に消費支出だけを考えている。

---

\* 本稿は科学研究費補助金若手研究Aによる研究成果の一部である

一方、近年の消費支出自体は実質減少しているものの、消費支出を細分化してみると、医療費、通信費、教養娯楽耐久消費財など実質増加している支出もある。これは社会の変化による消費行動の変化が起因である。つまり、各家計は消費できる範囲内で社会の変化や個人の興味に合わせ、支出傾向を変化させているといえることができるであろう。従来の消費関数では財政政策などにより可処分所得の変化が消費にどのように影響しているかを解明することが目的であるが、さらに消費行動を理解するには、消費を細分化し、それぞれの特徴を考慮した上で所得との関係を見ていく必要がある。

従来の消費関数のパラメータ推定には主に回帰分析が用いられている。これは消費支出とといった一つの目的変数を表す方程式を考えるモデルでは有効な統計的手法である。しかし、目的変数となる変数を細分化して複数個考える場合には回帰分析は適用できない。細分化した費目ごとに回帰分析を用いて消費関数を推定することもできるが、これらの費目は相互に関連しているため、あまり好ましくない。

本稿の目的は一般の家計において、消費行動が所得によりどのように変化しているのかをより詳しく分析するため、所得と消費を細分化した費目をそれぞれ考え、多変量解析の一つである正準相関分析を用いて解明することである。正準相関分析とは2群の相関が高くなるような重み付けを考え新たな変量を合成して作成する手法で、回帰分析と主成分分析の応用例として考えることができる。さらに、回帰分析は複数の説明変数（独立変数）を用いて単一の目的変数（従属変数）を表す推定式を作成するのに対して、正準相関分析では2つの変数群の関係を表す手法といえる。本論の分析の特徴は、さまざまな所得仮説に対する消費関数は消費として1つの値（例えば消費支出）しか考えていないのに対して、消費と所得をそれぞれ細分化して正準相関分析を行うことにより、消費と所得の関連性をさらに詳しく調べることにある。なお、本稿では総務省統計局作成の家計調査結果を用いることとする。

所得や消費を費目ごとに分類すれば、2つの関係に影響のある費目やない費目、恒常性の高い費目や変動性の高い費目などそれぞれ特徴があると思われる。本稿では所得と消費について以下の分類を行い、それぞれの費目の特徴を明らかにした上で、分析を行う。所得については経済学では可処分所得がよく用いられる。これは実収入から非消費支出を差し引いたものである。しかし非消費支出は一般的に実収入と関連が高いことから本稿では実収入に代替し、実収入を細分化した変数を用いて、各費目が消費支出にどのような影響を与えているのかを考えていく。消費については消費支出を用いる。家計調査では10大費目別の消費額が公表されていることから、これらの費目に分類して分析を行う。

本稿の構成は以下のとおりである。2章は平成15年度家計調査年報による現状の家計について考える。平成7年からの変化率から時系列的な変化を捉えたとともに、分析対象のデータに対して記述統計量を算出し、収入、支出の現状を探る。なお、変動性を調べるためには

それぞれの変数の平均値の大きさが異なるため、変動係数を用いて比較している。さらに従来の回帰分析に基づいた消費関数のパラメータ推定を行う。3章では本稿で主要となる正準相関分析の考え方と問題点について説明し、2章で作成した家計調査データを基に正準相関分析を行い、各正準変量の解釈について考える。4章では3章の結果についての考察を行う。

## 2 家計調査報告による分析

### 2.1 家計調査の調査と収支項目

本稿で用いるデータは平成15年1月から12月までの家計調査の勤労者世帯における都道府県庁所在地別のデータである。家計調査ではまず単身世帯と二人以上の世帯に分類され、さらに二人以上の世帯は勤労者世帯とそれ以外の世帯に分類されている。単身世帯の場合は二人以上の世帯に比べ、消費行動が異なり、勤労者以外の世帯に関しては収入があいまいなため、ここでは二人以上の勤労者世帯を調査対象としている。家計調査では約9000世帯に調査され、都道府県庁所在地別、年間収入別などに集計され、公表されている。このうち、都道府県庁所在地別(政令指定都市である川崎市、北九州市を含む49都市)平均値を日本の勤労者世帯のモデル的な世帯として、家計の収支項目をデータとした。平成15年の家計調査では二人以上の勤労者世帯は月平均で約4468世帯が調査されている。このうち、都道府県庁所在地別は約2941世帯が調査され、全勤労者世帯の約65.9%、全調査世帯の約32.7%の世帯が分析対象となる。ただし都市圏の世帯が標本対象となっているため、小都市や町村における消費と所得の関係については別の議論となるであろう。また、1年間を通して賞与など季節による変動性の高い収入項目があることや、本稿で主要となる正準相関分析は十分なデータ数を必要とすることから月別のデータを12ヶ月分用いて $49 \times 12 = 588$ 個のデータを用いた。

さて、家計調査報告では総収入を収入項目として約30個に分類され、集計されている。総収入は実収入とそれ以外、繰入金に大別される。経済において、所得として用いられるのは可処分所得が一般的である。可処分所得は実収入から非消費支出を差し引いたものである。本稿では消費関数の所得を可処分所得、実収入それぞれを用いて分析し、分析結果がほぼ同等であることを実証した後、さまざまな角度から分析を行うため、実収入を世帯主定期収入、世帯主臨時収入、世帯主賞与、世帯主以外の収入<sup>1</sup>、他の経常収入<sup>2</sup>、特別収入に分類し、6変数をデータとして扱う。一方、消費についても消費項目として約500個に分類され、集計されている。この中で家計の消費を最も表している消費支出について10大費目別(食料、住居、光熱・水道、家具・家事用品、被服及び履物、保健医療、交通・通信、教育、教養娯楽、そ

<sup>1</sup> 家計調査の分類における世帯主の配偶者の収入と世帯主以外の収入の合計額を世帯主以外の収入としている。

<sup>2</sup> 家計調査の分類における事業・内職収入と他の経常収入の合計額を他の経常収入としている。

表 1 主な家計収支の記述統計量

	世帯人員 (人)	年齢 (歳)	実収入 (万円)	可処分所得 (万円)	消費支出 (万円)	非消費支出 (万円)	黒字 (万円)
平均値	3.42	45.91	53.11	44.68	32.96	8.437	11.72
標準偏差	0.175	2.152	17.72	15.39	4.82	2.78	14.19
変動係数	0.051	0.047	0.334	0.344	0.146	0.329	1.211

他の消費支出)に分類した, 10 変数をデータとして扱う。

消費に関しては収入以外にも, 世帯人員や世帯主年齢, 職業などにも影響がある。本データは都市ごとに集計された世帯の平均値を用いていることにより, 世帯人員や世帯主年齢は平均化されているデータを用いている。実際に本稿で用いる 588 個のデータに対して世帯人員と世帯主年齢, 実収入, 可処分所得, 消費支出, 非消費支出, 黒字に対して記述統計量を計算したところ, 表 1 が得られた。表 1 の世帯人員と世帯主年齢をみると, それぞれ変動係数が極めて小さく, ほぼどのデータもこれらの影響力が除去されていることがわかる。

## 2.2 勤労者世帯における収支の時系列的な変化

本稿で用いるデータは平成 15 年のデータであるが, 近年の収入, 支出の時系列的な変化を見るために, まずは実質変化率を用いて近年の動向を考察しよう。表 2, 表 3 は平成 7 年から平成 15 年までの勤労者世帯における実収入の実質変化率と消費支出の実質変化率を費目ごとに細分化したものである<sup>3</sup>。まず収入の動向について表 2 をみると, 平成 10 年以降は主たる経常収入である定期収入や賞与をはじめ, ほとんどの収入が減少していることがわかる。これにより, 所得は費目に関係なく全体的に実質減少していることがわかる。

表 2 勤労者世帯における総収入の実質変化率

	平成 7	8	9	10	11	12	13	14	15
実収入	0.9	1.5	1.1	-1.8	-2.0	-1.5	-0.8	-1.2	-2.3
経常収入	1.0	1.6	1.2	-1.8	-2.1	-1.2	-0.9	-1.5	-2.1
定期収入	1.4	1.1	1.1	-1.3	-0.8	0.3	-1.3	-0.2	-0.7
臨時収入	-34.3	-1.2	2.3	-11.9	-8.8	-3.3	-6.4	3.3	-7.8
賞与	-2.7	2.8	0.9	-5.5	-7.4	-5.7	-2.5	-6.7	-4.3
配偶者収入	11.8	0.8	0.9	-1.1	-0.2	-3.2	-0.4	5.4	-3.3
他の世帯員収入	-8.9	0.1	3.6	0.1	-10.9	-6.5	-0.6	-10.5	-15.8
事業・内職収入	-20.3	1.3	-4.2	-17.3	3.9	10.3	16.3	-27.4	-12.8
他の経常収入	4.5	9.6	2.6	6.8	-2.4	-2.8	9.8	-11.7	-5.9
特別収入	0.1	-2.4	-0.4	0.8	0.9	-15.7	-0.4	10.4	-11.1
可処分所得	0.5	1.3	0.1	-0.9	-2.0	-1.4	-0.8	-1.5	-2.4

<sup>3</sup> 支出におけるその他の消費支出はこづかいや仕送り金が含まれているため, 実質値が計算できない。このため名目変化率を掲載している。

表3 消費支出の実質変化率

	平成7	8	9	10	11	12	13	14	15
消費支出	-0.7	0.6	0.1	-1.8	-1.7	-0.6	-0.8	-0.2	-1.2
食料	-1.9	-0.9	0.4	-1.0	-2.1	-2.2	-1.7	1.1	-2.2
調理食品	2.6	-0.8	5.9	1.3	-0.3	0.3	2.9	-0.2	0.7
住居	2.0	3.9	-3.6	-7.9	2.2	-3.8	2.8	-2.4	3.9
光熱・水道	1.9	2.3	-0.3	1.5	0.8	0.5	-0.8	-0.4	0.4
家具・家事用品	0.3	0.2	-0.8	-1.8	0.6	-4.5	4.8	-1.0	-0.9
被服及び履物	-3.5	-4.2	-3.1	-7.1	-0.9	-7.9	-3.7	-0.1	-0.5
保健医療	-1.8	4.9	0.8	-5.0	3.7	0.6	-1.7	-1.6	6.4
交通・通信	3.2	6.1	2.3	1.0	-1.5	7.2	1.6	-0.3	2.4
通信	6.1	8.7	12.9	5.2	10.4	11.1	16.4	7.1	6.6
教育	-5.4	-2.1	1.4	-3.9	-6.4	1.2	-4.1	-2.0	2.4
教養娯楽	-3.1	2.9	0.0	0.5	3.1	-3.2	2.2	1.1	-1.0
教養娯楽用耐久財	12.2	19.9	10.2	-1.3	19.0	7.4	22.1	11.9	3.9
その他の消費支出	-0.4	-1.2	1.7	-0.7	-5.0	-1.0	-3.5	-1.7	-5.7

消費支出，その他の消費支出は名目値

次に支出の動向について表3をみると，消費支出自体は平成7年以降実質減少をしていることがわかる。一方，細分化して各支出費目に着目すると，食料における調理食品費，交通・通信における通信費，教養娯楽における教養娯楽用耐久財などは堅調な伸びを見せている。調理食品費の増加は，共働きなど労働環境の変化が食変化につながっているため，増加傾向にあると思われる。通信費は携帯電話やインターネットの普及などIT産業の充実によるものである。教養娯楽用耐久財は主にテレビの需要によるものである。デジタル放送の開始など技術革新による買い替え行動により増加傾向にあると思われる。

このように収入は全体的に実質減少し，消費支出自体も実質減少しているものの，細分化すれば社会環境の変化により消費が移行していることにより増加している費目も多くあることがわかる。

## 2.3 平成15年の収入項目と支出項目の記述統計量

2.2節では時系列的な変化を見たが，本節では平成15年における勤労者世帯の各家計における近年の家計の収支の大きさや変動の大きさについて記述統計量から現状を把握していく。表4，表5は2.1節で分類した収入と消費支出の費目別の平均値，標準偏差，変動係数を計算した結果である<sup>4</sup>。まず収入について表4をみると，家計の収入は世帯主の「定期収入」が高いことがわかる。このことから家計の主だった収入源は世帯主の毎月定期的に得られる収

<sup>4</sup> 変動係数は都市別，月別も掲載している。都市別の変動係数は各都市における12か月分の平均値を算出し，変動係数を求めている。月別の変動係数については各月における49都市の平均値を算出し，変動係数を求めている。このように平均値を用いた変動係数であるため，各変数を比較する場合には有効ではあるが，都市別，月別の変動係数を比較することは意味をなさない。

表 4 主な収入項目の記述統計量

(円)

	平均	標準偏差	変動係数		
				(都市別)	(月別)
定期収入	359146.0	39267.1	0.1093	0.0953	0.0066
臨時収入	2069.7	3569.4	1.7246	0.6175	0.2511
賞与	69683.1	134236.3	1.9264	0.2260	1.9008
他の世帯員収入	62278.8	36031.1	0.5785	0.4418	0.2814
他の経常収入	24118.7	25520.1	1.0581	0.4982	0.7604
特別収入	13841.1	35662.6	2.5766	0.6858	0.6233

表 5 10 大費目別消費支出の記述統計量

(円)

	平均	標準偏差	変動係数		
				(都市別)	(月別)
食料	71141.3	7952.2	0.1118	0.0770	0.0657
住居	22777.5	11679.5	0.5128	0.3066	0.0720
光熱・水道	20836.7	3646.7	0.1750	0.0705	0.1435
家具・家事用品	10307.6	4179.6	0.4055	0.1649	0.1292
被服及び履物	15506.3	4792.0	0.3090	0.1475	0.1698
保健医療	11472.0	4106.4	0.3580	0.1540	0.0682
交通・通信	43147.9	18030.8	0.4179	0.1891	0.0915
教育	18736.4	10793.4	0.5761	0.2123	0.2925
教養娯楽	32407.6	8404.3	0.2593	0.1377	0.1160
その他	83272.3	23632.7	0.2838	0.1906	0.1078

入(実支出の約 67.6%)であることがわかる。次に変動係数を比較すると、「定期収入」,「世帯主以外の収入」が低いのに対して「特別収入」や「賞与」,「臨時収入」などは高い数値になっている。「特別収入」は遺産相続金などの受贈金や賭け事の配当金など定期性又は再現性のない特別な収入であることから特異な世帯があることによりこのような変動性が高い変数となっている。「賞与」はある一定の支給時期に一括して得られる収入であることから、月別変動が大きい。よって季節変動の大きい変数といえる。「臨時収入」は奨励金、報奨金、一時的な現金給与などその月に限って支給される収入であるため、「特別収入」と同様に変動性が高いものと思われる。よって恒常性の高い収入項目としては「定期収入」,「世帯主以外の収入」などが挙げられ、変動性の高い収入項目としては「特別収入」,「世帯主賞与」,「世帯主臨時収入」などが挙げられる。

次に消費支出について表 5 をみると、「その他の消費支出」,「食料費」,「交通・通信費」に関する支出が高いことがわかる。「その他の消費支出」は使途不明を含むこづかいや雑費、交際費が主な支出となっている。実際支出してはいるのだが、どの細目にも当てはまらない支出額も含まれている項目である。「食料費」に関しては外食も含む食費の合計額である。消費

から豊かさを測る場合にエンゲル係数が用いられるが、全体として消費支出の高い費目である。「交通・通信」に関しては、自動車の維持と通信に関するものが主な支出内容である。現在の自動車普及率の高さやIT産業の充実によるサービス供与により支出額が高くなっている。次に、変動係数を見ると、変動の高い費目は「教育」、「住居」となっている。「教育」については月別の変動係数が他の変数よりも高いことから、季節に影響している変動と考えられる。これは「教育」で主な支出となる学費は4月、10月に集中しているためである。「住居」の主な費目は、地代と設備修繕維持費である。「住居」の変動が大きい理由はこの設備修繕維持費が原因である。この費目は突発的に消費が行われるもので、高額であることから、変動係数が高くなる傾向がある。なお、都市別の変動係数のほうが月別よりも高くなっているが、これはサンプル数の違いによるものである。

## 2.4 消費支出に対する回帰分析

経済学において家計調査における一般的な消費関数は消費として消費支出、所得としては可処分所得と考えられる。上記データを用いて、目的変数を消費支出( $C$ , 実額)、説明変数を可処分所得( $Y$ , 実額)として回帰分析を行ったところ、

$$C = 274141.5 + 0.124 Y, \quad R^2 = 0.1573, \quad \bar{R}^2 = 0.1559 \quad (1)$$

と得られた。ここで  $R^2$  は決定係数、 $\bar{R}^2$  は自由度修正済決定係数を表す。また、平均消費性向は78.8%となった。決定係数が0.1573と低く、適合度が低い結果となった。このため、家計において可処分所得による消費の説明力は低いことを意味している。次に所得として実収入( $Y^*$ , 実額)と考えた場合の回帰分析を行ったところ、

$$C = 271387.6 + 0.1096 Y^*, \quad R^2 = 0.1626, \quad \bar{R}^2 = 0.1612 \quad (2)$$

と得られ、平均消費性向は66.0%となった。この二つの回帰式を比較すると、決定係数に大きな違いは見られず、どちらの変数を用いても分析精度には差がないことを表している。さらに決定係数がどちらも低いことから、絶対所得が消費に影響されているとする絶対所得仮説が信頼できないことを表している。

さらに、実収入を細分化し、説明変数を定期収入( $Y_1$ )、臨時収入( $Y_2$ )、賞与( $Y_3$ )、世帯主以外( $Y_4$ )、他の経常収入( $Y_5$ )、特別収入( $Y_6$ )、として、重回帰分析を行ったところ、

$$\begin{aligned} C = & 107484.4 + \frac{0.5491}{(0.4476)} Y_1 + \frac{0.3351}{(0.0248)} Y_2 + \frac{0.0320}{(0.0891)} Y_3 \\ & + \frac{0.2803}{(0.2097)} Y_4 + \frac{0.1191}{(0.0631)} Y_5 + \frac{0.1198}{(0.0887)} Y_6 \end{aligned} \quad (3)$$

表 6 可処分所得による 10 大費目別消費支出の回帰分析結果

	限界消費性向	切片	$R^2$	t 値
消費支出	0.1240	274141	0.1573	10.459
食費	0.0271	59034	0.2751	14.912
住居	0.0021	21820	0.0008	0.684
光熱水道	-0.0016	21572	0.0048	-1.687
家具家事	0.0062	7553	0.0515	5.642
被服履物	0.0072	12296	0.0532	5.741
保健医療	0.0041	9659	0.0231	3.724
交通通信	0.0052	40828	0.0020	1.074
教育	-0.0059	21386	0.0071	-2.054
教養娯楽	0.0171	24767	0.0981	7.983
その他	0.0628	55225	0.1671	10.844

を得た。ここで下段の数値は標準偏回帰係数を表している。修正済決定係数は  $\bar{R}^2=0.3095$  と (1), (2) のモデルより適合度のよいモデルが得られた。よって単回帰分析の結果よりも消費支出の現状を把握することができていると思われる。また標準偏回帰係数を見ると、「定期収入」と「世帯主以外の収入」が高いことから、この二つの収入項目が消費に影響を与えていることがわかる。統計的な見地からこの結果を解釈すると、これら二つの変数は実収入に対するウェイトが高く、変動性の低い変数である。よって恒常的に得られる収入が消費支出に大きく影響していると思われる。

ただしどのモデルも決定係数あるいは修正済決定係数の値が低いことからモデルの適合度はあまりよくないといえる。それは消費項目を消費支出という 1 つの変数で考えている点にある。表 4 のように収入を分類すると変動性の高い収入や低い収入など変数ごとに特性がある。この合計である実収入を説明変数とした(2)式よりも各変数を説明変数とした(3)式のほうがモデルの適合度はよい。これと同様に消費に関しても支出項目で変動性の高い支出や恒常的な支出があることから、消費支出を分類して解析することはさらに関連性の強い関係が得られることが期待される。そこで表 6 では消費支出を 10 変数に分類し、可処分所得による回帰分析を行ったところ、「食費」や「その他の支出」については消費支出に対する回帰式である(1)式と同じぐらいの適合度が得られたものの、他の変数はさらに低い結果となった。これは各費目に対して別々に分析したことにより、消費費目の関連性が無視されてしまったことが原因である。このことから次章では、消費費目の相関を考慮した分析を行うため、正準相関分析を用いて消費と所得の関係を考える。

### 3 正準相関分析を用いた所得と消費の関係

#### 3.1 正準相関分析とは

正準相関分析とは 2 つの変数間の相関構造を表す手法である。今、 $X = (x_1, \dots, x_p)'$ ,  $Y =$



$(y_1, \dots, y_q)'$  としよう。一般性を失わず  $p \leq q$  として、分散共分散行列は

$$\text{Var} \begin{pmatrix} X \\ Y \end{pmatrix} = \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \quad (4)$$

と定義されるものとする。ここで  $X, Y$  の一次結合  $f_1 = a_1'X, g_1 = b_1'Y$  を考える。正準相関分析とは合成変数  $f_1$  と  $g_1$  の相関を最大にするような重みベクトル  $a_1, b_1$  を定める手法で、最大となる相関係数を第1正準相関係数と呼ぶ。さらに  $k \leq p$  のとき、 $f_i, g_i (i=1, \dots, k-1)$  と独立となる  $X, Y$  の一次結合  $f_k = a_k'X, g_k = b_k'Y$  に対して、最大となる相関を第  $k$  正準相関係数と呼ぶ。実際には固有方程式

$$|\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} - \rho^2\Sigma_{xx}| = 0 \quad (5)$$

の  $i$  番目に大きい固有根の平方根が第  $i$  正準相関係数、それに対応する固有ベクトルが重みベクトル  $a_i$  となる。この重みベクトルを  $X$  の第  $i$  正準ベクトルと呼ぶ。 $Y$  の第  $i$  正準ベクトル  $b_i$  については

$$b_i = \frac{1}{\rho_i} \Sigma_{yy}^{-1} \Sigma_{yx} a_i \quad (6)$$

により得られる。また1次結合  $f_i$  および  $g_i$  を  $X, Y$  の第  $i$  正準変量と呼んでいる。標本が与えられたときの標本正準相関係数は  $\Sigma$  の代わりに標本分散共分散行列

$$S = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix} \quad (7)$$

を用いて定義される。固有方程式

$$|S_{xy}S_{yy}^{-1}S_{yx} - r^2S_{xx}| = 0 \quad (8)$$

の  $i$  番目に大きい固有根の平方根が第  $i$  標本正準相関係数となる。

正準相関分析は回帰分析と主成分分析の応用例として考えられる。もし  $Y$  が1変量の場合には正準相関係数は重相関係数と一致する。重相関係数は決定係数の平方根であることから  $\rho$  はモデルの適合度を表していることになる。重回帰分析では1つの目的変数  $Y$  に対して複数の説明変数  $X$  を組み合わせ、 $Y$  の予測式を推定する方法であるが、目的変数となりうる変数群に重み付けをして相互の変数群の関連を最大にするような重み付けを決定する方法が正準相関分析となる。

また、正準相関分析では、それぞれの相関が最大となるよう合成関数を生成させる。これは主成分分析と同じ座標軸を回転して生成することから、それぞれの正準変量は独立である

ことが保たれている。よって正準変量間の相関は

$$E(f_i f_j) = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases} \quad E(g_i g_j) = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases} \quad E(f_i g_j) = \begin{cases} \rho_i & (i=j) \\ 0 & (i \neq j) \end{cases} \quad (9)$$

となる。

次に正準相関係数の性質について考えよう。正準相関係数是不偏性の性質がなく  $O(1/\sqrt{n})$  のバイアスがある。標本正準相関係数に関する分布論は正規分布の場合であれば Muirhead (1982) などにある。また漸近分布論に関しては Seo et al. (1994) が一般論を展開している。ここで漸近分布論から得られる正規分布の場合の正準相関係数の期待値は

$$E(\sqrt{n}(r_i^2 - \rho_i^2)) = (1 - \rho_i^2) [p + q - 1 - 2\rho_i^2 + 2(1 - \rho_i^2) \sum_{j \neq i} \frac{\rho_j^2}{\rho_i^2 - \rho_j^2}] n^{-1/2} + O(n^{-2/3}) \quad (10)$$

となる (Siotani et al. (1985))。10式の  $O(1/\sqrt{n})$  のバイアスは次元数  $p, q$  に依存していることから、変数が増えればそれだけバイアスが大きくなる性質がある (A 参照)。よって標本データを用いる場合にはできるだけ変数を絞り込むことと可能であればデータ数を増やすことが必要となる。

このような観点から次元数に関する検定

$$H_0 : \rho_1 \geq \dots \geq \rho_{k+1} = \dots = \rho_p \quad (11)$$

が正準相関分析において重要な仮説検定法となる。これは逐次的に行われ、有効な次元数を決定する方法で、次元縮小の可能性を表している。このような検定問題に対しては検定統計量

$$T_k = -\{n - 1/2(p + q + 1)\} \log \Lambda_k \quad (12)$$

が帰無仮説の下で漸近的に自由度  $pq$  のカイ 2 乗分布に従うことにより行える<sup>5</sup>。ここで

$$\Lambda_k = (1 - r_{k+1}^2) \dots (1 - r_p^2) \quad (13)$$

<sup>5</sup> この仮説検定に関しては Lawley (1959), Seo et al. (1995) などさらに精度のよい近似を行える統計量を提案している。また Yamada (2003) ではこの仮説検定に対してノンパラメトリック検定を提案している。このほかにも正準相関係数に関する検定としては変量の有意性検定や個々の正準相関係数に対する検定法は有効な検定法である。特に高次の正準相関係数がいくつであるかは重要な問題となる。この問題について Yamada et al. (2001) でノンパラメトリック検定を提案している

とする。もし次元数が  $k$  であれば主成分分析と同様、 $k$  個まで次元が縮小できることから、 $k$  個まで変数選択が可能であることを示唆している。このような検定法を用いることにより変数の次元をできるだけ下げる努力をするべきである。ただし変数選択の方法などの変数の縮小に関する問題は多くの問題があり、まだ十分な議論がされていない。

次に、正準変量の解釈について考える。解釈を行う場合には正準ベクトルを用いて行う。ただし、正準ベクトルは変数の平均値に影響される。もし変数の単位が異なったり、平均値が極端に異なる場合には正準ベクトルの値では十分な解釈が行えない。そこで正準変量の解釈を行う場合は、正準変量と  $X, Y$  との相関係数を要素とする正準構造ベクトルを用いたほうがよい。これは主成分分析における主成分付加量と同等の性質がある。なお、正準変量の解釈には主成分分析の解釈と同様の難しさがある。

このように正準相関分析を行う場合は各変数が大きいと解釈の難しさとバイアス評価による誤差の両面において問題が生じる。よって解析する際には変数の検討をして、可能であれば変数の取捨、合成などを行い、できるだけ変数の次元を下げるのが望まれる。

### 3.2 家計調査における正準相関分析

2.3 節では消費支出に対する回帰分析を行い、消費と所得の関係について調べた。本節では消費支出について 10 変数、収入について 9 変数として正準相関分析を行い、さらに詳しく消費と所得の関係について分析する。

今、消費  $C$  を  $C_1, \dots, C_p$  と分類し、収入  $Y$  も  $Y_1, \dots, Y_q$  としよう。ここで

$$C = C_1 + \dots + C_p, \quad Y = Y_1 + \dots + Y_q$$

とする。各消費費目が収入に依存するとして

$$f_1(C_1, \dots, C_p) = f_2(Y_1, \dots, Y_q)$$

となる関数を推定する。従来の回帰分析では目的変数を単一の変数と考えるためこのような推定式は分析できない。そこで 3.1 節で紹介した正準相関分析を用いて当該問題を考えていくことにする。回帰分析による消費関数では、明示的に消費と所得の因果関係がはっきりしているモデルである。しかし、正準相関係数は相関係数の一種であることから、このような因果関係を表しているものではない。しかし、重みベクトルを用いた正準変量を合成して新たな変数を得ることにより、従来よりもさらに強い関係を導くことができる。正準相関分析を用いた消費と収入の関係について調べる利点は、収入群とそれに影響のある消費群を特徴づけ、新たな因果関係を導出することにある。この点については次章で考察する。

さて、2.1 節で定義したデータを用いて分析したところ、表 7 の結果が得られた。この結果

表 7 家計調査における正準相関分析結果

	正準相関係数	寄与率	累積寄与率
$\rho_1$	0.6904	38.1%	38.1%
$\rho_2$	0.4099	22.6%	60.6%
$\rho_3$	0.3551	19.6%	80.2%
$\rho_4$	0.1649	9.1%	89.3%
$\rho_5$	0.1094	6.0%	95.3%
$\rho_6$	0.0844	4.7%	100.0%

から第一正準相関係数は 0.6904 であり、重回帰分析を行った(3)式のモデルの重相関係数が 0.5627 であることからさらに関連性の高い関係を見出していることになる。さらに第二、第三正準相関係数もそれぞれ 0.4099, 0.3551 を得た。よって、第一正準相関係数とは別の角度から消費と所得の関係を低いながらも表しているといえる。なお、第三正準相関係数までの累積寄与率が 80.2%であることから、約 8 割の関係が見出せていることになる。

次に正準構造ベクトルを用いてそれぞれの正準変量の解釈について考えよう。収入の第一正準構造ベクトルを見ると、「定期収入」、「賞与」、「世帯主以外の収入」の関連性が高いことがわかる。これらの変数は収入額が高いという共通点があり、これらの合計は収入総額の約 92.5%を占めていて、どれも勤め先収入である。よって収入の第一正準変量は「雇用労働所得」と解釈できる。消費の第一正準構造ベクトルをみると「食料」、「その他の消費支出」、「教養娯楽」の関連性が高いことがわかる。支出の中でもこれらの費目の支出が収入群の第一正準変量である雇用労働所得に影響していることがわかる。このことから収入が増えることにより、食事が豊かになり、その他、教養娯楽といった余暇活動を楽しむといった行動が活発になることが伺える。次に、収入の第二正準構造ベクトルを見ると、「世帯主以外の収入」、「他の経常収入」、「特別収入」といった世帯主の収入以外の収入が正の値、「賞与」、「定期収入」、「臨時収入」といった世帯主の収入が負の値となっていることから、世帯主の収入と世帯主収入以外の二極型の正準変量である。よってこの正準変量は収入源に関する正準変量と考えることができる。消費の第二正準構造ベクトルをみると「その他」、「光熱・水道」といった費目が正の高い値、「被服履物」、「保健医療」、「住居」などが負の値となっている。このことから世帯主の収入以外の収入が大きい世帯では「その他の消費支出」、「光熱・水道」の消費が高いといえる。最後に、収入の第三正準構造ベクトルを見ると、「定期収入」と「賞与」がそれぞれ正負の高い数値となっている。よってこれは「定期収入」と「賞与」の二極型の正準変量である。消費の第三正準構造ベクトルをみると「教育」、「光熱・水道」、「教養娯楽」、「交通・通信」、「被服履物」などが高い正の値となっている。このことから賞与の割合が高い時期にはこれらの消費があまりされないことがわかる。

図1：収入の第一正準構造ベクトル

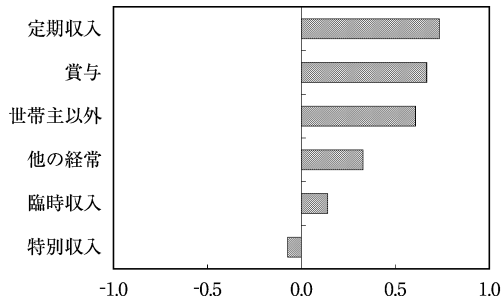


図2：支出の第一正準構造ベクトル

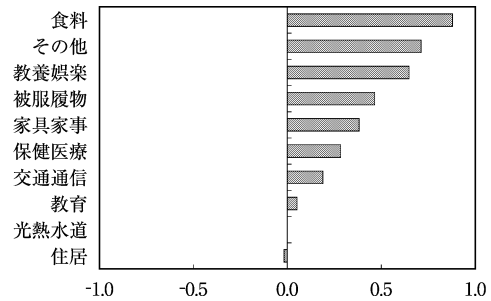


図3：収入の第二正準構造ベクトル

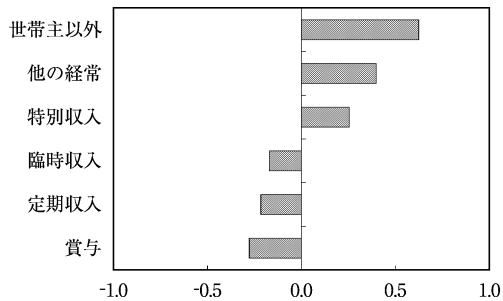


図4：支出の第二正準構造ベクトル

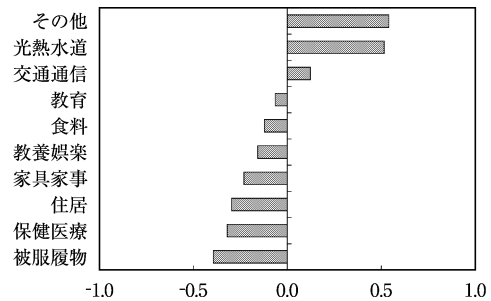


図5：収入の第三正準構造ベクトル

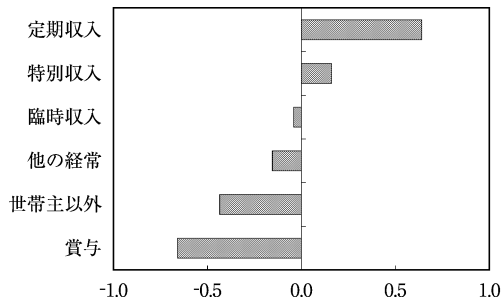
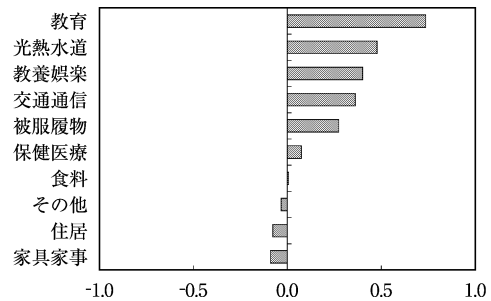


図6：支出の第三正準構造ベクトル



## 4 考 察

本稿では平成15年度の家計調査の勤労者世帯について、消費と所得についての関係をより明確にするために正準相関分析を用いて分析した。分析結果は以下の通りである。

まず、第一正準相関係数は0.6904と得られた。これは消費支出に関する重回帰式である(3)のモデルの重回帰係数よりも高いため、従来の回帰分析による消費関数よりも精度の高い関連性が見出されたといえるだろう。収入の第一正準変量は雇用労働所得とした。今回の分析では勤労者世帯に限定しているため、実収入のほとんどが雇用労働所得であったことが、「そ

の他の経常収入」などの影響力を弱める原因となったと思われる。よって第一正準変量は実収入、あるいは可処分所得と解釈することもできるだろう。次に支出の第一正準変量は「食料」、「その他の支出」、「教養娯楽」との関係が高い変数であることがわかった。これらの費目の支出額は節約や浪費など生活に合わせ可変的な支出額である。また消費が高くなることにより、食事が豊かになり、余暇活動などが充実すると思われる。よってこれらの消費額は「生活の豊かさに関する支出」と考えることができるだろう。一方、相関の低かった「教育」、「光熱・水道」などは生活水準を維持する上で必要となる支出額である。消費支出  $C$  を

$$C = C_1 + C_2$$

（ここで  $C_1$  は生活の豊かさに関する支出（食料、教養娯楽、その他の合計額）、 $C_2$  は生活水準を維持するための支出）として、 $C_1$  に対して可処分所得を説明変数として回帰分析を行ったところ、

$$C_1 = 139026.0 + 0.107 Y, \quad R^2 = 0.2670, \quad \bar{R}^2 = 0.2658 \quad (14)$$

を得た。一方、 $C_2$  に対しても同様の回帰分析を行ったところ、

$$C_2 = 1135115.5 + 0.0172 Y, \quad R^2 = 0.0090, \quad \bar{R}^2 = 0.0073 \quad (15)$$

を得た。この結果から  $C_1$  に対しては可処分所得を説明変数とした(1)式のモデルよりも適合度がよい結果となっている一方で、 $C_2$  は決定係数があまりにも低いため収入に依存しない消費であることがわかる。このことから所得は家計の消費支出全体に影響があるという考えよりも生活の豊かさに対する消費支出に影響しているといえるであろう。

次に、第二、第三正準相関係数はそれぞれ 0.4099, 0.3551 と得られた。これらは弱い関係ながらも副次的に所得と消費の相関構造が見出されたといえるであろう。収入の第二正準変量は収入源に関する変数であることがわかった。特に収入の第二正準変量是有業人員との相関が高く、有業人員を表す変数といえる。よって家計の収入は有業人員に影響され、有業人員の差により消費行動も変化していることがわかる。また、支出の第二正準変量は「その他の消費支出」や「光熱水道」との関連性の高い変数であることがわかった。このことから、有業人員が増えれば消費の中で「その他の消費支出」や「光熱水道」費が増えていることがわかる。まず、「その他の消費支出」との関係であるが、有業人員と「その他の消費支出」をさらに細分化した費目との相関を見ると、他の変数に比べれば高い相関があるため、「その他の消費支出」全体が有業人員に影響されているということであろう。また、有業人員が多いということは世帯人員にも関係してくる。一般的に世帯人員が多い世帯は光熱・水道費に関する支出が高いことから、支出の第二正準変量との相関が高いことが推察される。収入の第

三正準変量は収入の変動性に関する変量であることがわかった。収入は賞与により季節変化の大きい変量である。よってその季節性を表した変数といえる。また、支出の第三正準変量は「教育」や「光熱水道」、「教養娯楽」、「交通・通信」などとの関連性の高い変数であることがわかった。これらの費目も季節変動の高い変数といえるが、所得に応じてではなく、生活サイクルに依存して季節変動する費目といえる。「教育」費の季節的変動要因は主に授業料である。授業料は年度始めあるいは2期に分割して支出される費目である。よって4月、10月に支出額が高くなっている。「光熱・水道」費の季節的変動要因は気温である。特に光熱費は冬季に多く支出される傾向がある。「教養娯楽」費の季節的変動要因は教養娯楽サービスによるものである。これは休暇に影響されるため、長期休暇が得られる時期に比較的支出される傾向がある。「交通・通信」費の季節的変動要因は自動車の維持費である。特に自動車税の納入時期に関係しているものと思われる。このように所得の変動とは別個の季節変動による消費の変化がわかった。

以上のように、所得と消費の関係について考えた場合、本稿で定義した「豊かさに関する消費」が所得に影響していることがわかり、従来の消費関数よりも適合度の高い関係が得られた。また、副次的に有業人員により収入構造の変化、またそれとともに消費行動の変化があることがわかった。収入、消費いずれも季節変動の大きなものであるが、消費の中で所得の変動とは別の消費サイクルのある費目を見出すことができた。このことから近年の家計収支動向が把握できたものと思われる<sup>6</sup>。

## 5 おわりに

本稿では正準相関分析の実用性を実証するために経済学で基本となる消費行動の解明について分析した。従来の経済学におけるモデル推定のほとんどは回帰分析によるもので、一つの変数を表すモデルが考えられている。回帰式を独立して複数個推定するという考え方もできるが、目的変数間にも相関がある問題が多いため本研究のように正準相関分析を用いたほうがより説明力が高くなる。このように正準相関分析は回帰分析の応用例であることから、経済学でもさまざまな分野で必要な分析方法と考えられる。

分析データは都道府県庁所在都市別データを用いた。これはその都市の調査対象の平均値である。これらの平均値をモデル的な世帯の収支項目として分析したが、年齢層、世帯人員、ローンの有無、子供の有無などさまざまな異質の世帯があると思われる。また、所得処分モデルでは消費や貯蓄は所得のほかに物価や利率、資産なども影響されていると考えられて

<sup>6</sup> 本稿の分析データは平成15年の家計調査を用いたが、平成14年の同じデータを用いても同様の結果が得られている。

いる。本分析では、単年度データを扱ったため、物価や利子率については一定であるという条件の下で、消費には影響されないとして分析を行った。資産についてはライフサイクル仮説あるいは恒常所得仮説の下では重要な変数であろう。資産を表す貯蓄については毎月家計の収支状況と同時に報告され、集計されているが、公表は3ヵ月毎の四半期データのみである。よって今回の分析には加えることができなかったが、この点も考慮して分析するべきであろう。現在、秘匿問題から個表データの取得は難しく、本研究のように集計データによる分析せざるを得ないのが現状であるが、より精密な分析を行うためには個表データを用いたほうがよいと思われる。

本研究では消費と所得に関する分析を行ったが、主成分分析における解釈と同様の難しさやバイアスの影響など分析方法自体にも問題点がある。これらの問題はすべて変量の次元数に依存するといってよいだろう。このような観点から、さらに分析精度を高めるためには分析者が変数の吟味を十分に検討する必要があるとともに、変数選択に関する数理統計学の進歩が必要となるであろう。

また、現在行われている消費推計は今期の所得や資産だけでなく将来の所得も加味したモデルが採用されている。今回分析したデータは今期の消費と所得だけを用いているため、これらのモデルよりも簡単なモデルを用いて分析を行った。しかし、本分析を将来の所得も含めたモデルに应用することは可能であると思われる。現在の消費推計は消費の絶対量に対する分析が主流である。しかし本分析のように消費を細分化した消費ベクトルを用いて消費を推計することは社会の変化などによる消費の趣向性も考慮に入れられるため、重要な分析であると思われる。この点に関しては今後の研究課題である。

## 謝 辞

本論文を作成するに当たり、多くの助言をいただいた播磨谷浩三氏に深く感謝する。

## A 数値計算による標本正準相関係数と正準相関分析における問題点

標本正準相関係数がどのくらいのバイアスがあるかを数値計算により示そう。サンプル数を  $N=50, 100, 200, 500$  として、多変量正規分布(M.N.),  $\epsilon$  汚染正規分布( $\epsilon=.1, \sigma^2=3$ (C.N.)), 多変量  $t$  分布 ( $\nu=12$  (M.T.)) の3つの楕円分布族についての数値計算を行う。 $X, Y$  をそれぞれ5次元 ( $p=q=5$ ) の変数として分散共分散行列(4)を簡単のため

$$\Sigma = \begin{pmatrix} I & P \\ P' & I \end{pmatrix}, \quad P = [\text{diag}(.70, .30, .05); 0] \quad (16)$$

としよう。ここで  $I$  は単位行列を表す。このとき、各サンプル数、分布に対する正準相関係



表A 数値計算による標本正準相関係数

N		$\rho_1 = .70$	$\rho_2 = .30$	$\rho_3 = .05$	$\rho_4 = .00$	$\rho_5 = .00$
50	M.N.	0.761	0.489	0.322	0.181	0.057
	C.N.	0.761	0.490	0.322	0.181	0.057
	M.T	0.769	0.509	0.340	0.192	0.060
100	M.N.	0.729	0.397	0.239	0.131	0.041
	C.N.	0.729	0.397	0.239	0.131	0.041
	M.T	0.734	0.411	0.253	0.139	0.043
200	M.N.	0.714	0.346	0.177	0.095	0.029
	C.N.	0.714	0.346	0.177	0.095	0.029
	M.T	0.717	0.355	0.189	0.102	0.031
500	M.N.	0.706	0.318	0.119	0.063	0.019
	C.N.	0.706	0.318	0.119	0.062	0.019
	M.T	0.707	0.321	0.127	0.067	0.021

M.N. : 多変量正規分布

C.N. :  $\epsilon$  汚染正規分布 ( $\epsilon = .1, \sigma^2 = 3$ )M.T. : 多変量  $t$  分布 ( $\nu = 12$ )

数は表Aとなる。この結果から正準相関係数は高々次元数が5次元であってもサンプル数が十分ないと分析結果として信頼のおけないものとなる。 $N=50$ の時は、どの分布においても母正準相関係数とはかなりの差があることがわかる。通常の分析では $N=100$ ぐらいサンプル数があれば十分な近似が行えるため、分析精度が保たれていると言われているが、正準相関分析の場合は $N=200$ であってもバイアスが大きいことがわかる。これは次元数が増えれば、それだけ分析精度が低くなるという「次元の呪い」(curse of dimension)という問題にも関連している。正準相関分析の場合には2群の変数に対して分析するため、 $p+q$ 次元変量を扱う。よって、その他の多変量解析法よりも多くの次元数を必要とすることからさらに分析精度が低くなっている。さらに $\rho_i$ の値が小さければバイアスの及ぼす影響が大きくなる。このバイアスは $\rho_i$ 以外の正準相関係数や高次のキュムラントにも影響している<sup>7</sup>が、次元数に関する部分 $(1-\rho_i^2)(p+q-1-2\rho_i^2)n^{-1}$ の影響力が大きいことがわかっていてる。

正準相関係数は行列式(5)あるいは(8)により計算される。この行列は逆行列を用いた表現となっている。よって $X$ 群や $Y$ 群に高い相関を持つ変数がある場合には重回帰分析における多重共線性と同様の問題から解が不安定になる。よってこのような変数がある場合には変数を再構築する必要がある。さらに重解を含む( $\rho_i = \rho_{i+1}$ )場合にも問題となる。これは行列式から固有値、固有ベクトルを計算する際に重解があると正しく計算されないためである。正準相関分析で結果を吟味する場合にはこれらの問題点を理解した上で行う必要があるだろう。

<sup>7</sup> 正規分布の場合、高次のキュムラントはすべて0であるため、(10)では高次のキュムラントはない簡単な表現がされているが、一般論を展開する場合はさらに複雑な表現がされる。

参考文献

- [1] 林 文夫 (1986) 「恒常所得仮説の拡張とその検証」『経済分析（経済企画庁経済研究所）』 101.
- [2] Hotelling, H. (1936) Relations between two sets of variates, *Biometrika*, 28, 321-377.
- [3] 羽森茂之 (1996) 『消費者行動と日本の資産市場』 東洋経済新報社
- [4] Muirhead, R. J.; Waternaux, C. (1980) Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika*, 67, 31-43.
- [5] Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York.
- [6] Lawley, D. N. (1959) Tests of significance in canonical analysis, *Biometrika*, 46, 59-66.
- [7] Seo, T., Kanda, T.; Fujikoshi, Y. (1994) The effects on the distributions of sample canonical correlations under nonnormality, *Commun. Statist.-Theory Met.*, 23, 2615-2628.
- [8] Seo, T., Kanda, T. and Fujikoshi, Y. (1995). The effects of nonnormality on tests for dimensionarity in canonical correlation and MANOVA models. *J. Multivariate Anal.*, 52, 325-337.
- [9] Siotani, M., Hayakawa, T.; Fujikoshi, Y. (1985) *Modern Multivariate Analysis: A Graduate Course and Handbook*, Amrican Sciences Press, Ohio.
- [10] 総務省統計局編 (2002) 「平成 15 年度家計調査年報」 日本統計協会.
- [11] Yamada, T.; Sugiyama, T. (2001) On the permutation test for canonical correlation coefficients, *Proceeding in 53rd ISI*, 135-136
- [12] 山田智哉 (2003) 正準相関分析における次元数に関するノンパラメトリック検定 2003 年度統計関連学会連合大会報告

(やまだ ともや 統計科学・経済統計学専攻)

(2004 年 11 月 1 日 受理)